# An Empirical Study of the Usefulness of SARFIMA Models in Energy Science

Leila Sakhabakhsh[1], Masoud Yarmohammadi[2]
[1] M.Sc. of Statistics, Tehran North Branch, Islamic Azad University, Tehran, Iran
[2]Department of Statistics, Faculty of Science, Payame Noor University, Tehran, Iran.
[1]leila.sakhabakhsh@yahoo.com; [2]masyar@pnu.ac.ir

***Abstract*-In this paper the specification of long memory has been studied using monthly data in consumption of petroleum products of OECD during (Jan 1994 to Mar 2010). As monthly data of consumption of petroleum products of OECD seems to be non-stationary and shows periodic behavior, the data is fit with SARIMA and SARFIMA models, and is estimated the parameters using conditional sum of squares method. The results indicate the appropriate model is SARFIMA (2, 1, 0) (0, 0.473, 0)$_{12}$ which is used to predict the consumption rate of petroleum products of OECD till the end of 2013.**

***Keywords-Long Memory, SARFIMA Model, SARIMA Model, Petroleum, OECD***

## I. INTRODUCTION

At the beginning of the 20th century, petroleum was a minor resource used to manufacture lubricants and fuel for kerosene and oil lamps. One hundred years later it had become the preeminent energy source for world.

Total world energy usage rises from 495 quadrillion British thermal units (Btu) in 2007 to 590 quadrillion Btu in 2020 and 739 quadrillion Btu in 2035 (Fig. 1).

The global economic recession that began in 2008 and continued into 2009 has had a profound impact on world energy demand in the near term. Total world marketed energy consumption contracted by 1.2 percent in 2008 and by an estimated 2.2 percent in 2009, as manufacturing and consumer demand for goods and services declined. Although the recession appears to have ended, the pace of recovery has been uneven so far, with China and India leading and Japan and the European Union member countries lagging. In the reference case, as the economic situation improves, most nations return to the economic growth paths that were anticipated before the recession began.

The most rapid growth in energy demand from 2007 to 2035 occurs in nations outside the organization for economic co-operation and development (non- OECD nations). Total non-OECD energy consumption increases by 84 percent in the Reference case, compared with a 14-percent increase in energy use among OECD countries [1].

OECD is an international economic organization of 34 countries founded in 1961 to stimulate economic progress and world trade. It is a forum of countries committed to democracy and the market economy, providing a platform to compare policy experiences, seek answers to common problems, identify good practices, and co-ordinate domestic and international policies of its members.

In this paper we intend to forecast consumption of petroleum products of OECD using SARIMA and SARFIMA models to construct an adequate model.

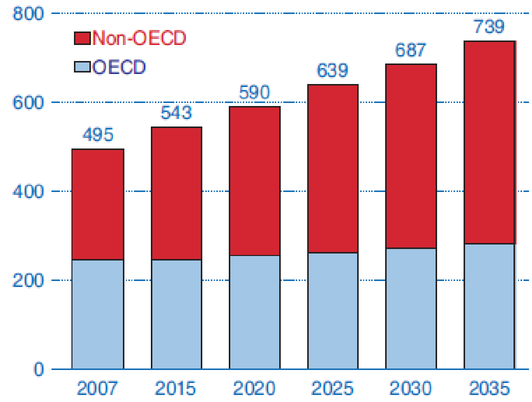



Fig. 1 World marketed energy consumption, 2007-2035 (quadrillion Btu) [1]

The recent finance and economic literature has recognized the importance of long memory in analyzing time series data. A long memory can be characterized by its autocorrelation function that decays at a hyperbolic rate. Such a decay rate is much slower than that of the time series, which has short memory. Traditional Box-Jenkins models describing short memory, such as AR (p), MA (q), ARMA (p, q), and ARIMA (p, d, q) cannot describe long memory precisely. A set of models has been established to overcome this difficulty, and the most famous one is the autoregressive fractionally integrated moving average (ARFIMA or ARFIMA (p, d, q)) model. ARFIMA model was established by Granjer and Joyeux [2]. An overall review about long memory and ARFIMA model was model by Baillie [3]. In many practical applications researchers have found time series exhibiting both long memory and cyclical behavior. For instance, this phenomenon occurs in revenues series, inflation rates, monetary aggregates, and gross national product series. Consequently, several statistical methodologies have proposed to model this type of data including the Gegenbauer autoregressive moving average processes (GARMA), k-factor GARMA processes, and seasonal autoregressive fractionally integrated moving average (SARFIMA) models. The GARMA model was first suggested by Hosking [4] and later studied by Gray et al [5] and Chung [6]. Other extension of the GARMA process is the k-factor GARMA models proposed by Giratis and Leipus [7] and Woodward et al [8]. This paper investigates a special case of the k-factor GARMA model, which is considered by Porter – Hudak [9] and naturally extends the seasonally integrated autoregressive moving average (SARIMA) model of Box and Jenkins [10]. Katayama [11] examined the asymptotic properties of the estimators and test statistics in SARFIMA models. Mostafaei and Sakhabakhsh [12] used SARIMA and SARFIMA models to construct an adequate model for Iraqi oil production. There are several methods for estimating the parameters in time series models. In this paper, we estimate the parameters using

conditional sum of squares (CSS) method and testing procedures using residual autocorrelations such as the Lagrange multiplier (LM) test are shown.

This paper is organized of follows: Section II gives some definitions and properties for the ARFIMA and SARFIMA processes. Section III illustrates the use of the SARIMA and SARFIMA models and section IV presents our final conclusions.

## II. MATERIALS & METHODS

### A. Arfima Model

Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a white noise process with zero mean and variance $\sigma_\varepsilon^2 > 0$, and B the backward-shift operator, i.e., $B^k(X_t) = X_{t-k}$ . If $\{X_t\}_{t\in\mathbb{Z}}$ is a linear process satisfying

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_{t}, \qquad t\in\mathbb{Z}. \tag{1}$$

Where d∈ ( -0.5, 0.5), $\Phi(.)$ , $\Theta(.)$ are polynomials of degree p and q, respectively, given by

$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \text{ }, \Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

Where $\phi_i, 1 \le i \le p$, $\theta_j, 1 \le j \le q$, are real constants, than $\{X_t\}_{t\in\mathbb{Z}}$ is called general fractional differentiation ARFIMA (p, d, q) process, where d is the degree or fractional differentiation parameter. If d ∈ (-0.5;0.5), then $\{X_t\}_{t\in\mathbb{Z}}$ is a stationary, and an invertible process. The most important characteristic of an ARFIMA (p, d, q) process is the property of long dependence, when d ∈ (0.0;0.5), short dependence, when d=0, and intermediate dependence, when d ∈ (-0.5;0.0).

### B. SARFIMA (P, D, Q) (P, D, Q)S Processes

In many practical situation time series exhibit a periodic pattern. We shall consider the SARFIMA (p, d, q) (P, D, Q)s process, which is an extension of the ARFIMA process.

Definition 1. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary stochastic process with spectral density function $f_x(.)$.suppose there exists a real number $b \in (0,1)$ , a constant $C_f > 0$ and one frequency $G \in [0,\pi]$ (or a finite number of frequencies ) such that

$$f_x(\omega) \sim c_f|\omega - G|^{-b}, \text{ when } \omega \to G.$$

Then, $\{X_t\}_{t\in\mathbb{Z}}$ is a long memory process.

Remark 1. In Definition 1, when $b \in (-1,0)$, we say that the process $\{X_t\}_{t\in\mathbb{Z}}$ has the intermediate dependence property [13].

Definition 2. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stochastic process given by the expression

$$\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D(x_t - \mu) = \theta(L)\Theta(L^s)\varepsilon_t, \ t \in \mathbb{Z} \tag{2}$$

Where μ is the mean of the process, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process with zero mean and variance $\sigma_\varepsilon^2 = \mathbb{E}({\varepsilon_t}^2)$, $s \in \mathbb{N}$ is the seasonal period, L is the backward-shift operator, that is $L^{sk}(X_t) = X_{t-sk}$ , $\nabla_s^D = (1-L^s)^D$ is the seasonal difference operator, $\phi(.)$, $\theta(.)$,$\Phi(.)$, and $\Theta(.)$ are the polynomials of degrees p, q, P, and Q, respectively, defined by

$$\phi(L) = \textstyle\sum_{i=0}^{p}(-\phi_i)L^i, \qquad \theta(L) = \sum_{j=0}^{q}(-\theta_j)L^j,$$

$$\Phi(L) = \textstyle\sum_{k=0}^{P}(-\Phi_k)L^k, \qquad \Theta(L) = \sum_{l=0}^{Q}(-\Theta_l)L^l, \tag{3}$$

Where, $\phi_i, 1 \le i \le p, \theta_j, 1 \le j \le q, \Phi_k, 1 \le k \le P$ , and $\Theta_l, 1 \le l \le Q$ are constants and $\phi_0 = \Phi_0 = -1 = \theta_0 = \Theta_0$.

Then, $\{X_t\}_{t\in\mathbb{Z}}$ is a seasonal fractionally integrated ARMA process with periods, denoted by SARFIMA (p, d, q) (P, D, Q)s, where d and D are the order of differencing and the seasonal differencing respectively.

Theorem 1. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a SARFIMA (p, d, q) (P, D, Q)s process given by the expression (2), with zero mean and seasonal period $s \in \mathbb{N}$ [14].

Suppose $\phi(z)\Phi(z^s) = 0$ and $\theta(z)\Theta(z^s) = 0$ have no common zeroes. Then, the following are true

(i) The process $\{X_t\}_{t\in\mathbb{Z}}$ is stationary if d + D < 0.5, D < 0.5 and $\phi(z)\Phi(z^s) \ne 0$ , for $|z| \le 1$.

(ii) The stationary process $\{X_t\}_{t\in\mathbb{Z}}$ has a long memory property if 0 < d + D < 0.5, 0 < D < 0.5 and $\phi(z)\Phi(z^s) \ne 0$ , for $|z| \le 1$.

(iii) The stationary process $\{X_t\}_{t\in\mathbb{Z}}$ has an intermediate memory property if -0.5 < d + D < 0, -0.5 < D < 0 and $\phi(z)\Phi(z^s) \ne 0$ , for $|z| \le 1$ .

### C. Conditional Sum of Squares Method

There are several methods for estimating the parameters in time series models. In this paper, we implement the CSS method to estimate the SARIMA and SARFIMA models of consumption of petroleum products of OECD. This method is equivalent to the full Maximum Likelihood Estimator (MLE) under quite general conditional homoskedastic distributions. A description of the properties of the CSS estimator and its finite sample performance is presented in Chung and Baillie [15].

### D. Lagrange Multiplier Test

This section discusses testing for the integration order, namely, the LM test, which draws on LM tests for the integration order of the ARFIMA model by Robinson [16,17], Agiakloglou and Newbold [18], and Tanaka [19]. For the purpose of practical implementation, Godfrey's [20] LM approach is also used. For the SARFIMA model, we consider the testing problem of the null hypothesis H0: SARFIMA (p, d, q) (P, D, Q)s against the alternative:

HA, 1: SARFIMA(p, d + α0, q) (P, D, Q)s

or HA, 2: SARFIMA (p, d, q) (P, D + αs, Q)s

The assumed null model is obtained by imposing the restriction $\alpha_0(\alpha_s) = 0$ and the alternatives are $\alpha_0(\alpha_s) > 0$ or $\alpha_0(\alpha_s) < 0$.

We get the p-values for testing the integration order corresponding to tests.

## III. EMPIRICAL STUDY

### A. The Data

In this study we will use the monthly consumption of petroleum products of OECD during (Jan 1994 to Mar 2010). The data are obtained from the energy information administration of the U.S. department of energy. Fig. 2 displays the data of consumption rate of petroleum products of OECD, $\{x_t\}$.

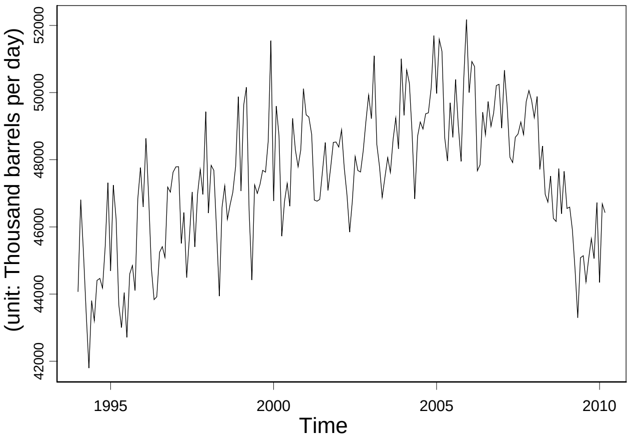
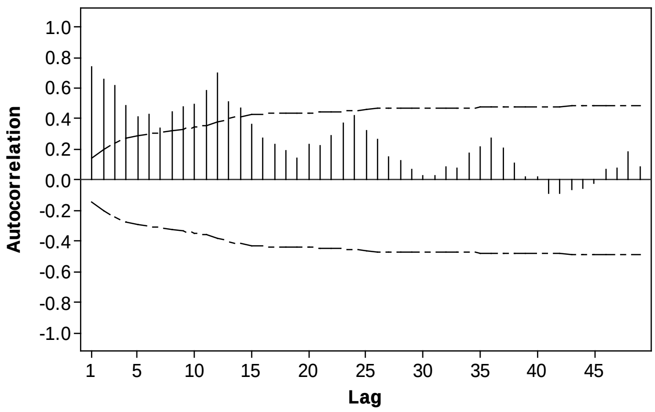


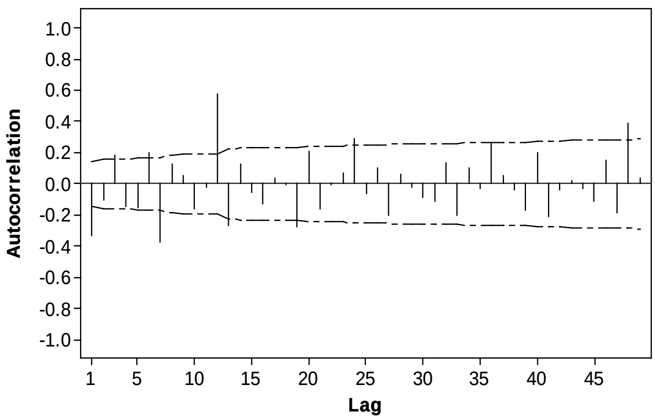

Fig. 2 Time plot of consumption rate of petroleum products in OECD

(with 5% significance limits for the autocorrelations)




Series: A

(with 5% significance limits for the autocorrelations)




Series: B

Fig. 3 The sample autocorrelation function (ACF) of the transformed series, where A is $\{x_t\}$, B is $\{(1-B)x_t\}$. Dotted lines are approximate 95% confidence limits of the ACF of the white noise random variable.

Fig. 3 displays the autocorrelation function (ACF) of the transformed data, $\{x_t\}$. The ACF decays very slowly and exhibits cyclical behavior.

As it is shown in figure 3, this time series is non-stationary and doesn't require seasonal differencing.

*B. Model Selection*

To search for the best representation of this data, we first fitted differenced data $y_t = (1-B)x_t$ by the CSS method, where we used a sample mean of $\{y_t\}$, $\bar{y}$ as an estimator of $E[y_t] = \mu$, and set s = 12. AIC criteria are also used under the assumption of normality [21].

Fitting SARFIMA models or SARIMA models is limited to having SARMA parameters with $0 \le p, q, P, Q \le 3$, and where the total number of estimated SARFIMA parameters (d, D, SARMA parameters, and $\sigma^2$) is less than 4. The total number of models is 70. As mentioned earlier, in addition to SARIMA models, SARFIMA models are fitted as well, because we intend to determine if the consumption of petroleum products in OECD have long memory. From among these estimation results, we selected models in terms of AIC. All calculations were made using S-PLUS. Table 1 shows the best five models in terms of AIC model selection with estimators. ID denotes the model identification within 70 models. NE indicates the corresponding parameter is not estimated and is set to be 0. The numbers in parentheses in the column of AIC denote the ranking of models in terms of AI

TableI SUMMARY OF AIC MODEL SELECTION AND ESTIMATES

| ID | AIC | d | D | | $\phi_2$ | $\theta_1$ | $\Phi_1$ | $\Theta_1$ |
|---|---|---|---|---|---|---|---|---|
| 53 | (1)2996.8 | NE | 0.47 | -0.6 | -0.5 | NE | NE | NE |
| 13 | (2)3010.6 | NE | NE | -0.6 | -0.5 | NE | 0.7 | NE |
| 63 | (3)3010.9 | NE | 0.42 | NE | NE | 0.7 | NE | -0.25 |
| 59 | (4)3013.0 | NE | 0.46 | NE | NE | 0.7 | 0.1 | NE |
| 54 | (5)3013.1 | NE | 0.55 | NE | NE | 0.7 | NE | NE |

TableII P-VALUES OF TESTING FOR $A_0 = A_S = 0$ OF THE SARFIMA MODELS

| Model | Alternative hypotheses | | |
|---|---|---|---|
| | $\alpha_0 > 0$, $\alpha_s = 0$ | $\alpha_0 = 0$, $\alpha_s > 0$ | $\alpha_0 \neq 0$ , $\alpha_s \neq 0$ |
| SARFIMA (2, $\alpha_0$, 0) (0, $\alpha_s$, 0) | 0.5226 | 0 | 0 |
| SARFIMA (2, $\alpha_0$, 0) (1, $\alpha_s$, 0) | 0.1724 | 0.6046 | 0.1861 |
| SARFIMA (0, $\alpha_0$, 1) (0, $\alpha_s$, 1) | 0.4492 | 0 | 8.21×10-11 |
| SARFIMA (0, $\alpha_0$, 1) (1, $\alpha_s$, 0) | 0.4775 | 0.0305 | 0 |
| SARFIMA (0, $\alpha_0$, 1) (0, $\alpha_s$, 0) | 0.1744 | 0 | 0 |

From the table 1 the SARFIMA (2, 0, 0) (0, 0.473, 0)12 model (model ID: 53) is the best model in terms of AIC among the 70 model candidates. From theorem 1, the process $\{y_t\}$ has long memory property. The appropriate model for $y_t$ is SARFIMA (2, 0, 0) (0, 0.473, 0)12 model therefore the best model for $x_t$ is SARFIMA (2, 1, 0) (0, 0.473, 0)12 model.

Table 2 shows the p-values for testing the integration order corresponding to the best five models using the LM test statistics.

In this table, models ID 53, ID 63, ID 59 and ID 54 correspond to some models in alternative hypotheses of the first, third, forth and fifth rows of SARFIMA models, and models ID 13 correspond to null hypotheses of the second rows of SARFIMA models.

Our findings as follows: (i) results for SARFIMA (2, $\alpha_0$, 0) (0, $\alpha_s$, 0), SARFIMA (0, $\alpha_0$, 1) (0, $\alpha_s$, 1), SARFIMA (0, $\alpha_0$, 1) (1, $\alpha_s$, 0) and SARFIMA (0, $\alpha_0$, 1) (0, $\alpha_s$, 0) support the estimation of d or D for models ID 53, ID 63, ID 59 and ID 54. (ii) Except for SARFIMA (2, $\alpha_0$, 0) (1, $\alpha_s$, 0), results for

SARFIMA models show small p-values for the alternative $\alpha_0 = 0$, $\alpha_s > 0$ and $\alpha_0 \neq 0$, $\alpha_s \neq 0$.

### C. Forecasting

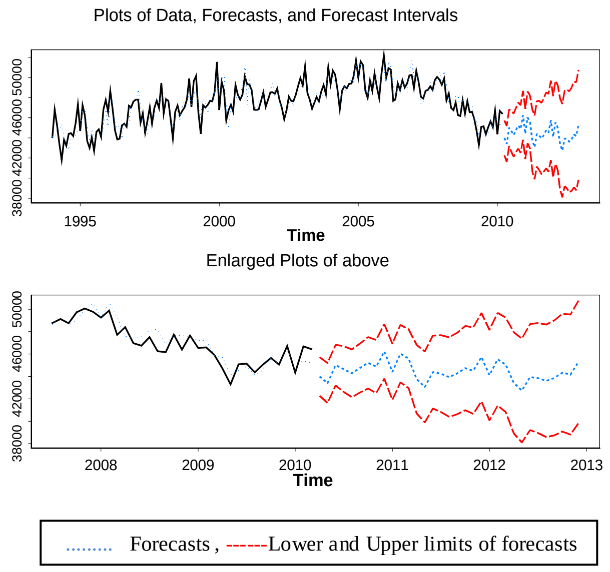



Fig. 4 Prediction plot of consumption rate of petroleum products of OECD (2010-2013)

The appropriate model is SARFIMA (2, 1, 0) (0, 0.473, 0)12 model which is used to predict the consumption rate of petroleum products of OECD till the end of 2013 as shown in Fig. 4. The results of in-sample and out-sample forecasts of the SARFIMA model are shown in table 3 and table 4.

TableIII OUT-SAMPLE FORECASTS FOR THE SARFIMA (2, 1, 0) (0, 0.473, 0)12 MODEL

| Date | Prediction | lowerCL | upperCL |
|---|---|---|---|
| 2010-04 | 43988.68 | 42263.02 | 45714.34 |
| 2010-05 | 43404.08 | 41614.76 | 45193.40 |
| 2010-06 | 44998.61 | 43182.91 | 46814.31 |
| 2010-07 | 44645.39 | 42594.60 | 46696.19 |
| 2010-08 | 44271.22 | 42134.14 | 46408.30 |
| 2010-09 | 44750.01 | 42561.50 | 46938.53 |
| 2010-10 | 45211.13 | 42915.38 | 47506.87 |
| 2010-11 | 44873.19 | 42500.42 | 47245.97 |
| 2010-12 | 46229.30 | 43797.88 | 48660.73 |
| 2011-01 | 44401.51 | 41896.88 | 46906.13 |
| 2011-02 | 46022.50 | 43452.26 | 48592.75 |
| 2011-03 | 45572.97 | 42945.66 | 48200.29 |
| 2011-04 | 43752.29 | 40694.10 | 46810.47 |
| 2011-05 | 43063.07 | 39893.88 | 46233.26 |
| 2011-06 | 44388.68 | 41138.55 | 47638.80 |
| 2011-07 | 44239.21 | 40808.55 | 47669.87 |
| 2011-08 | 43937.36 | 40392.99 | 47481.72 |
| 2011-09 | 44258.06 | 40622.80 | 47893.32 |
| 2011-10 | 44742.57 | 40988.97 | 48496.18 |
| 2011-11 | 44511.53 | 40656.58 | 48366.48 |
| 2011-12 | 45720.70 | 41776.63 | 49664.77 |
| 2012-01 | 44104.34 | 40065.01 | 48143.67 |
| 2012-02 | 45531.54 | 41402.56 | 49660.51 |
| 2012-03 | 45031.46 | 40819.12 | 49243.77 |

lowerCL: lower confidence limits of forecasts

upperCL: upper confidence limits of forecasts

TableIV IN-SAMPLE FORECASTS FOR THE SARFIMA (2, 1, 0) (0, 0.473, 0)12 MODEL

| Date | Actual | Forecasts | Error |
|---|---|---|---|
| 2009-01 | 46549 | 47211.08 | -662.08 |
| 2009-02 | 46589 | 47234.33 | -645.33 |
| 2009-03 | 45915 | 45896.04 | 18.96 |
| 2009-04 | 44724 | 45702.56 | -978.56 |
| 2009-05 | 43297 | 44139.81 | -842.81 |
| 2009-06 | 45083 | 44513.45 | 569.55 |
| 2009-07 | 45143 | 44879.68 | 263.32 |
| 2009-08 | 44367 | 44128.88 | 238.12 |
| 2009-09 | 45053 | 44551.60 | 501.4 |
| 2009-10 | 45649 | 45687.76 | -38.76 |
| 2009-11 | 45059 | 44988.31 | 70.69 |
| 2009-12 | 46725 | 46262.83 | 462.17 |

As it is shown in fig. 4, consumption rate of petroleum products of OECD is going to have descending trend in the future.

## IV. CONCLUSIONS

This paper has examined a seasonal long memory process, denoted as the SARFIMA model. As an illustration of the use of SARFIMA model, we considered monthly consumption of petroleum products of OECD. We fitted the data with SARIMA and SARFIMA models, and estimated the parameters using CSS method.

The results indicated the appropriate model was SARFIMA (2, 1, 0) (0, 0.473, 0)12 model which was used to predict the data. On the basis, we conclude that the SARFIMA model is effective and can be usefully employed as a substitute for the SARIMA model when fitting consumption rate of petroleum products of OECD.

## REFERENCES

[1] International Energy Outlook 2010. www.eia.gov/oiaf/ieo/index.html. July 2010

[2] C.W. Granjer and R. Joyeaux, " An introduction to long-memory time series models and fractional differencing," The Journal of Time Series Analysis, vol. 1, 1980, pp. 15-29.

[3] R.T. Baillie, " Long memory processes and fractional integration in econometrics," vol. 73, 1996, pp. 5-59

[4] J.R.M. Hosking, " Fractional differencing," Biometrika, vol. 68, 1981, pp. 165-76.

[5] H.L. Gray, N.F. zhang and W.A. Woodward, " On generalized fractional processes," J.Time Ser.Anal, vol. 10, 1989, pp. 233-57.

[6] C.F. Chung, " Estimating long memory process," J. Econometrics, Forthcoming, 1996.

[7] L. Giraitis, R. Leipus, " A generalized fractionally differencing approach in long memory modeling," Lithuanian mathematical Journal, vol. 35, 1995, pp. 65-81

[8] W.A. Woodward, Q.C. Cheng and H.L. Gray, " A k-factor long memory model," Journal of time series Analysis, vol. 19, 1998, pp. 485-504.

[9] Porter-Hudak, " An application of the seasonal fractionally differenced model to the monetary aggregates," J-Amer, Statist. Assoc, vol. 84(410), 1990, pp. 338-344.

[10] G.E.P. Box and G.M. Jenkins, " Time series analysis forecasting and control," 2nd ed. Holden-Day, San Francisco, 1976.

[11] N. Katayama, " Seasonally and fractionally differenced time series," Hitotsubashi Journal of Economics, vol. 48, 2007, pp. 25-55.

[12] H. Mostafaei and L. Sakhabakhsh, " Using SARFIMA and SARIMA models to study and predict the Iraqi oil production," Journal of basic and applied scientific research, vol. 1(10), 2011, pp. 1715-1723.

[13] P. Doukhan, G. Oppenheim and M.S. Taqqu, (2003). " Theory and applications of long-range dependence," Birkheuser, Boston, 2003.

[14] C. Bisognin and R.C. Lopes, " Properties of seasonal long memory processes," Mathematical and Computer Modeling, vol. 49, 2009, pp. 1837-1851.

[15] C.F. Chung and R.T. Baillie, " Small Sample Bias in Conditional Sum-of-Squares Estimators of Fractionally Integrated ARMA Models," Empirical Economics, vol. 18, 1993, pp. 791-806.

[16] P.M. Robinson, " Testing for strong serial correlation and dynamic conditional Heroskedasicity in multiple regressions," Journal of econometrics, vol. 47,1991, pp. 67-84.

[17] P.M. Robinson, " Efficient tests of non-stationary hypotheses" Journal of American statistical Associantion, vol. 89, 1994, pp. 1420-1437.

[18] C. Agiakloglou and P. Newbold , " Lagrange Multiplier Tests for fractional difference," Journal of Time Series Analyisis, vol. 15, 1994, pp. 253-262.

[19] K. Tanaka, " The non-stationary fractional unit root," Econometric theory, vol. 15, 1999, pp. 549-582.

[20] L.G. Godfrey, " Testing the adequacy of a time series model," Biometrika, vol. 66, 1979, pp. 67-72.

[21] P.J. Brockwell and Davis, Time series: Theory and Methods. Springer, New York, 1991.